Capítulo 8 - Testes de hipóteses

8.1 Introdução

Nos capítulos anteriores vimos como estimar um parâmetro desconhecido a partir de uma amostra (obtendo estimativas pontuais e intervalos de confiança para o parâmetro).

Muitas situações práticas têm uma natureza diferente, requerendo que em função dos valores observados se tomem decisões acerca dos parâmetros (ou de outros aspectos) da população.

Exemplos:

1) Máquina de encher pacotes de açúcar. O peso de cada pacote deve ser $\approx 8g$ (isto é, $\mu = 8$). Será que a máquina está a funcionar correctamente?

2) O pKa tabelado para o ácido orto-hidroxibenzóico é 2.81 ($\mu_y = 2.81$). Será que o valor calculado experimentalmente está de acordo com o valor tabelado, isto é, a eventual diferença entre os dois será devida apenas a erros acidentais ou existirá também erro sistemático?

Definição: Uma hipótese estatística é uma afirmação acerca dos parâmetros de uma ou mais populações (testes paramétricos) ou acerca da distribuição da população (testes de ajustamento).

Vamos estudar em primeiro lugar os testes paramétricos.

Exemplo 1) (cont.): temos duas hipóteses: a máquina funciona correctamente ($\mu = 8$) ou a máquina não funciona correctamente ($\mu \neq 8$):

$$H_0: \mu = 8 \quad \text{versus} \quad H_1: \mu \neq 8$$

(hipótese nula) (hipótese alternativa)

Hipótese simples: é especificado apenas um valor para o parâmetro.

Hipótese composta: é especificado mais de um valor para o parâmetro.

Vamos considerar sempre H_0 como hipótese simples.

A hipótese alternativa (H_1) é, em geral, uma das três seguintes:

$H_1: \mu \neq 8$ hipótese alternativa bilateral

$H_1: \mu > 8$ hipótese alternativa unilateral (superior)

$H_1: \mu < 8$ hipótese alternativa unilateral (inferior)

Nota: os valores especificados nas hipóteses não devem ter nada a ver com valores observados na amostra.

Definição: Teste de hipóteses é um procedimento que conduz a uma decisão acerca das hipóteses (com base numa amostra).

Exemplo 1) (cont.): X - v.a. que representa o peso de um pacote de açúcar, $E(X) = \mu$, $V(X) = \sigma^2$.

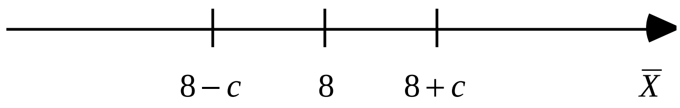
$$H_0: \mu = 8 \quad \text{versus} \quad H_1: \mu \neq 8$$

Dispomos de uma amostra de 10 observações:

$(X_1, \ldots, X_{10})$ (a.a.)

Faz sentido decidir com base em $\overline{X}$, aceitando H_0 se $\overline{X}$ estiver próxima de 8 e rejeitando H_0 se $\overline{X}$ estiver longe de 8.

região crítica	região de aceitação	região crítica
"Aceitar" H_1	"Aceitar" H_0	"Aceitar" H_1
Rejeitar H_0	Não rejeitar H_0	Rejeitar H_0



Região crítica: $\overline{X} < 8 - c$ ou $\overline{X} > 8 + c$

Aos pontos de fronteira chamam-se valores críticos.

Tipos de erro:

Decisão:	Situação: H_0 é verdadeira	H_0 é falsa
"Aceitar" H_0	não há erro	**erro do tipo II**
Rejeitar H_0	**erro do tipo I**	não há erro

$$\alpha = P(\text{erro do tipo I}) = \\ = P\left(\text{Rejeitar } H_0 \middle| H_0 \text{ é verdadeira}\right)$$

A α chama-se nível de significância.

$$\beta = P(\text{erro do tipo II}) = \\ = P\left(\text{"Aceitar" } H_0 \middle| H_0 \text{ é falsa}\right)$$

Voltando ao exemplo, vamos admitir que fazíamos $c = 0.5$ e que $\sigma = 1$ e $n = 10$.

A região crítica é: $\overline{X} < 7.5$ ou $\overline{X} > 8.5$.

Supondo que $X \sim N(\mu, 1)$ então $\overline{X} \sim N\left(\mu, \frac{1}{10}\right)$

$$\alpha = P\left(\overline{X} < 7.5 \text{ ou } \overline{X} > 8.5 \middle| \mu = 8\right) = \\ = \Phi\left(\frac{7.5-8}{\sqrt{0.1}}\right) + 1 - \Phi\left(\frac{8.5-8}{\sqrt{0.1}}\right) = 0.1142$$

Se aumentarmos *n*, mantendo os valores críticos, α diminui.

Quanto a β, não vamos ter um único valor mas uma função, ou seja, para cada μ de H_1 podemos calcular um valor $\beta(\mu)$. Por exemplo, para $\mu = 9$:

$$\beta(9) = P\left(\text{aceitar } H_0 \middle| \mu = 9\right) = \\ = P\left(7.5 \le \overline{X} \le 8.5 \middle| \mu = 9\right) = \\ = \Phi\left(\frac{8.5-9}{\sqrt{0.1}}\right) - \Phi\left(\frac{7.5-9}{\sqrt{0.1}}\right) = 0.0571$$

$$\beta(10) = P\left(\text{aceitar } H_0 \middle| \mu = 10\right) = \\ = P\left(7.5 \le \overline{X} \le 8.5 \middle| \mu = 10\right) = \\ = \Phi\left(\frac{8.5-10}{\sqrt{0.1}}\right) - \Phi\left(\frac{7.5-10}{\sqrt{0.1}}\right) \approx 0$$

Por simetria $\beta(7) = \beta(9)$ e $\beta(6) = \beta(10)$

Se mudarmos a região crítica, com *n* fixo:

Se *c* diminuir, α aumenta e, para cada μ, $\beta(\mu)$ diminui.

Se *c* aumentar, α diminui e, para cada μ, $\beta(\mu)$ aumenta.

É mais fácil controlar α do que controlar β (que depende de μ em H_1). Logo:

- rejeitar H_0 é uma conclusão "forte".
- "aceitar" H_0 é uma conclusão "fraca". Em vez de dizer "aceita-se H_0" é preferível dizer "não se rejeita H_0", ou "não há evidência suficiente para rejeitar H_0".

Definição: Chama-se <u>potência do teste</u> à probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira ($=1-\beta$).

No exemplo, a potência do teste quando $\mu = 9$ é $1-0.0571 = 0.9429$, ou seja, se a verdadeira média for 9, a diferença em relação a 8 será detectada 94.29% das vezes.

Como decidir entre alternativa unilateral ou bilateral?

I) $H_0: \mu = 8$ versus $H_1: \mu > 8$

Região crítica: $\overline{X} > 8 + c$

<u>Ponto de vista do fabricante!</u>

Quando rejeitar H_0 pára a produção para afinar a máquina.

©Ana Pires, IST, Outubro de 2000 **Probabilidades, Erros e Estatística**

II) $H_0: \mu = 8$ versus $H_1: \mu < 8$

Região crítica: $\overline{X} < 8 - c$

<u>Ponto de vista do consumidor!</u>

Quando rejeitar H_0 não aceita a encomenda

III) $H_0: \mu = 8$ versus $H_1: \mu \neq 8$

Região crítica: $\overline{X} < 8 - c$ ou $\overline{X} > 8 + c$

<u>Compromisso entre os dois!</u>

©Ana Pires, IST, Outubro de 2000 **Probabilidades, Erros e Estatística**

Procedimento Geral dos Testes de hipóteses

1. Pelo contexto do problema identificar o parâmetro de interesse
2. Especificar a hipótese nula
3. Especificar uma hipótese alternativa apropriada
4. Escolher o nível de significância, α
5. Escolher uma estatística de teste adequada
6. Fixar a região crítica do teste
7. Recolher uma amostra e calcular o valor observado da estatística de teste
8. Decidir sobre a rejeição ou não de H_0

©Ana Pires, IST, Outubro de 2000 **Probabilidades, Erros e Estatística**

8.2 Testes de hipóteses para a média, variância conhecida

X população tal que:

$E(X) = \mu$ (desconhecido)

$V(X) = \sigma^2$ (conhecido)

$(X_1, \ldots, X_n)$ a. a. de dimensão *n*

$X \sim N(\mu, \sigma^2)$ ou *X* qq com *n* grande.

Teste de $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$

Sabemos já que, quando H_0 é verdadeira

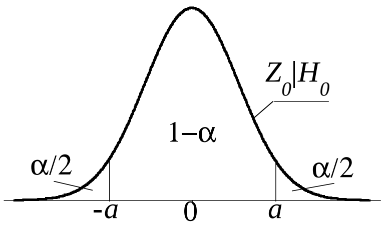
$$\overline{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \text{ou} \quad \overline{X} \underset{a}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

É conveniente estandardizar e usar como

estatística de teste: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

Quando H_0 é verdadeira $Z_0 \sim N(0,1)$

A região crítica deve ser bilateral porque H_1 é bilateral:



R.C.: $Z_0 < -a$ ou $Z_0 > a$ com a: $P(Z > a) = \dfrac{\alpha}{2}$

(recordar que $\alpha = P\left(\text{Rejeitar } H_0 \middle| H_0 \text{ é verdadeira}\right)$)

Seja $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ o valor observado da estatística de teste. Então

rejeita-se H_0 se $z_0 < -a$ ou $z_0 > a$

e não se rejeita H_0 se $-a \le z_0 \le a$

Estas regras podem ser expressas em termos de $\bar{x}$

rejeita-se H_0 se $\bar{x} < \mu_0 - a\dfrac{\sigma}{\sqrt{n}}$ ou $\bar{x} > \mu_0 + a\dfrac{\sigma}{\sqrt{n}}$

e não se rejeita H_0 se $\mu_0 - a\dfrac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + a\dfrac{\sigma}{\sqrt{n}}$

Exemplo 1) (cont.): X - v.a. que representa o peso de um pacote de açúcar (supõe-se que $X \sim N(\mu, 1)$). A máquina está afinada quando $\mu = 8$. Numa amostra de 25 pacotes (recolhida aleatoriamente) observou-se $\bar{x} = 8.5$.

Quer-se saber se, ao nível de significância de 5%, se pode afirmar que a máquina continua afinada.

$H_0: \mu = 8$ versus $H_1: \mu \ne 8$ (1. 2. e 3.)

Nível de significância = 5% (4.)

Estatística de teste: $Z_0 = \dfrac{\overline{X} - 8}{1 / \sqrt{25}}$ (5.)

$\alpha = 0.05 \Rightarrow a = 1.96$ donde

R.C.: $Z_0 < -1.96$ ou $Z_0 > 1.96$ (6.)

Com $\bar{x} = 8.5$ obtém-se $z_0 = \dfrac{8.5 - 8}{1/\sqrt{25}} = 2.5$ (7.)

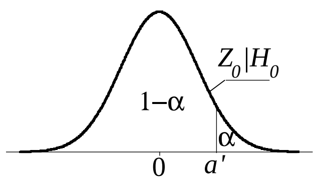
Como $z_0 > 1.96$ rejeita-se H_0, ou seja, existe evidência (ao nível de significância considerado) de que a máquina está desafinada.

Alternativas unilaterais

1) Se fosse $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$

estatística de teste: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

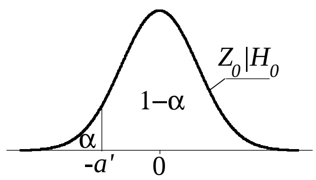
R.C.: $Z_0 > a'$ onde a': $P(Z > a') = \alpha$



2) Se fosse $H_0: \mu = \mu_0$ versus $H_1: \mu < \mu_0$

estatística de teste: a mesma

R.C.: $Z_0 < -a'$ onde a': $P(Z > a') = \alpha$

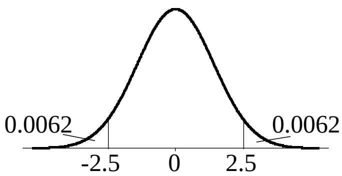


Outro método: valor-p

Em vez de fixar α, determinar a região crítica e, em seguida, verificar se o valor observado pertence à região crítica, pode olhar-se directamente para o valor observado da estatística de teste e determinar para que nível de significância a decisão muda.

Definição: Dado o valor observado da estatística de teste, o **valor-p** (*p-value*) é o maior nível de significância que levaria à não rejeição da hipótese nula (ou o menor que levaria à rejeição).

No exemplo, $z_0 = 2.5$, para este valor H_0 não é rejeitada se $\alpha \le 2[1-\Phi(2.5)] = 0.0124$, ou seja, $p = 0.0124$.



<u>Quanto mais baixo for o valor-p maior é a evidência contra a hipótese nula.</u>

Relação entre intervalos de confiança e testes de hipóteses:

Parâmetro desconhecido θ.

I.C. a $100\times(1-\alpha)\%$ para $\theta=[l,u]$, baseado numa dada amostra e v. a. fulcral, então a mesma amostra leva à rejeição de

$$H_0:\theta=\theta_0 \quad \text{contra} \quad H_1:\theta\neq\theta_0,$$

ao nível de significância α, se e só se $\theta_0 \notin [l,u]$

ou à não rejeição de H_0 se e só se $\theta_0 \in [l,u]$

Nota: é necessário que a v.a fulcral e a estatística de teste sejam da mesma forma.

Vamos ver que isto é verdade para o teste que estamos a estudar (teste para a média com variância conhecida):

$$H_0:\mu=\mu_0 \text{ versus } H_1:\mu\neq\mu_0$$

Não se rejeita H_0, ao nível de significância α, se e só se

$$\mu_0 - a\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + a\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$

$$\Leftrightarrow \bar{x} - a\frac{\sigma}{\sqrt{n}} \le \mu_0 \le \bar{x} + a\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$

$$\Leftrightarrow \mu_0 \in I.C._{100\times(1-\alpha)\%}(\mu)$$

No exemplo, $n=25$, $\bar{x}=8.5$, $\sigma=1$, I.C. a 95% ($\alpha=0.05$) $\Rightarrow$ $a=1.96$

$I.C._{95\%}(\mu)=[8.108;8.892]$, como $\mu_0=8$ não pertence ao I.C., rejeita-se $H_0:\mu=8$ (contra $H_1:\mu\neq 8$) ao nível $\alpha=5\%$.

Nota: o teste que acabámos de estudar é aplicável com σ^2 desconhecida (substituída por S^2) desde que a dimensão da amostra seja grande ($n>30$).

8.3 Testes de hipóteses sobre a igualdade de duas médias, variâncias conhecidas

X_1, população 1, com $E(X_1)=\mu_1$ e $V(X_1)=\sigma_1^2$ (conhecida)

X_2, população 2, com $E(X_2)=\mu_2$ e $V(X_2)=\sigma_2^2$ (conhecida)

(X_1 e X_2 independentes)

a. a. da população 1 $(X_{11},\ldots,X_{1n_1})$ com média $\bar{X}_1$

a. a. da população 2 $(X_{21},\ldots,X_{2n_2})$ com média $\bar{X}_2$

(e a a.a. $(X_{11},\ldots,X_{1n_1})$ é independente da a.a. $(X_{21},\ldots,X_{2n_2})$)

Queremos testar

$H_0 : \mu_1 = \mu_2$ contra uma das alternativas

$H_1 : \mu_1 \neq \mu_2$ (bilateral) ou

$H_1 : \mu_1 > \mu_2$ (unilateral superior) ou

$H_1 : \mu_1 < \mu_2$ (unilateral inferior)

já sabemos que

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Então, quando H_0 é verdadeira ($\mu_1 - \mu_2 = 0$)

$$Z_0 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Daqui em diante é tudo semelhante ao caso anterior, ou seja, dadas as amostras concretas calcula-se

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Com $H_1 : \mu_1 \neq \mu_2$, rejeita-se H_0 para o nível de significância α se

$$z_0 < -a \quad \text{ou} \quad z_0 > a \quad \text{com} \quad a : P(Z > a) = \frac{\alpha}{2}$$

etc.

Nota: este teste é válido para variâncias desconhecidas (substituídas por S_1^2 e S_2^2) desde que $n_1 > 30$ e $n_2 > 30$.

8.4 Testes de hipóteses para a média de uma população normal, variância desconhecida

Se $n < 30$ só é possível efectuar testes para a média se for possível assumir que $X \sim N(\mu, \sigma^2)$.

Nesse caso para testar

$H_0 : \mu = \mu_0$ contra uma das alternativas

$H_1 : \mu \neq \mu_0$ (bilateral) ou

$H_1 : \mu > \mu_0$ (unilateral superior) ou

$H_1 : \mu < \mu_0$ (unilateral inferior)

usa-se a estatística de teste $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

Quando H_0 é verdadeira $T_0 \sim t_{n-1}$

Então para $H_1 : \mu \neq \mu_0$, rejeita-se H_0 ao nível de significância α se

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -a \quad \text{ou} \quad t_0 > a$$

com $a : P(T_{n-1} > a) = \dfrac{\alpha}{2}$

etc.

Nota: Para os testes em que a estatística de teste tem distribuição normal o valor-p é fácil de determinar. Para outras distribuições (*t* e chi-quadrado) esse valor só pode ser obtido usando um programa de computador ou em certas calculadoras. Recorrendo às tabelas o melhor que se consegue é obter um intervalo que contém (de certeza) o valor-p.

Exemplo: Determinação da constante de acidez do ácido orto-hidroxibenzóico. O valor tabelado para pKa é 2.81. Queremos saber se o valor determinado experimentalmente <u>está de acordo com o valor tabelado</u> (devendo-se as eventuais diferenças apenas a erros acidentais) ou se, pelo contrário, o valor obtido <u>não está de acordo com o valor tabelado</u> (podendo então afirmar-se que ocorreu algum erro sistemático). Ou seja, em termos de testes de hipóteses e sendo Y a v.a. que representa um valor de pKa determinado experimentalmente, queremos testar

$$H_0: \mu_Y = 2.81 \quad \text{contra} \quad H_1: \mu_Y \neq 2.81$$

Admitindo que $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$

<u>Relatório 1</u>: Como já determinámos no Cap. 7 (pág. 22)

$$I.C._{95\%}\left(\mu_Y\right) = [2.8117; 2.8768]$$

fazendo uso da analogia descrita, como $2.81 \notin I.C._{95\%}\left(\mu_Y\right)$, podemos concluir que, ao nível de significância de 5%, rejeitamos H_0.

<u>Relatório 2</u>: A conclusão é semelhante

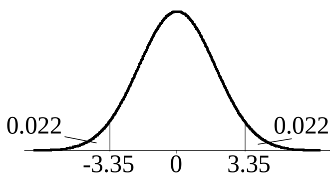
$$I.C._{95\%}\left(\mu_Y\right) = [3.0877; 3.1249]$$

No entanto é visível uma diferença grande entre os dois relatórios que pode ser melhor avaliada calculando o valor-p para cada teste.

Para isso precisamos calcular o valor observado da estatística de teste.

<u>Relatório 1</u>:

$$t_0 = \frac{\bar{y} - 2.81}{s_y / \sqrt{4}} = \frac{2.8443 - 2.81}{0.020468 / \sqrt{4}} = 3.35$$



valor - p = 0.044

(pelas tabelas: $0.01 < p < 0.05$)

<u>Relatório 2</u>:

$$t_0 = \frac{3.1063 - 2.81}{0.014946 / \sqrt{5}} = 44.33 \qquad \text{valor} - \text{p} \approx 0$$

Ao nível de significância de 5% conclui-se que há erro sistemático nos dois relatórios, usando o valor-p conclui-se que há muito maior evidência a favor da existência de erro sistemático no relatório 2 do que no relatório 1.

8.5 Testes de hipóteses sobre a igualdade das médias de duas populações normais, variâncias desconhecidas

Exemplo: Comparação entre os resultados obtidos pelos dois grupos na determinação da constante de acidez do ácido orto-hidroxibenzóico (vamos usar os dados relativos a pKa).

Teste de $H_0: \mu_1 = \mu_2$ contra $H_1: \mu_1 \neq \mu_2$

Relatório 1	Relatório 2
$\bar{y}_1 = 2.844275$	$\bar{y}_2 = 3.1063$
$s_{y1} = 0.020468$	$s_{y2} = 0.014946$
$n_1 = 4$	$n_2 = 5$

Admitimos que (hipóteses de trabalho):

- A primeira amostra é uma concretização de uma a.a. de uma população $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$;
- A segunda amostra é uma concretização de uma a.a. de uma população $Y_2 \sim N\left(\mu_2, \sigma_2^2\right)$;
- Y_1 e Y_2 são independentes;
- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (parece razoável porque s_{y1}^2 e s_{y2}^2 são da mesma ordem de grandeza).

$$s_p = \sqrt{\frac{3 \times 0.020468^2 + 4 \times 0.014996^2}{7}} = 0.017527$$

$$n_1 + n_2 - 2 = 7$$

Estatística de teste:

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad \text{sob } H_0 \quad T_0 \sim t_{n_1+n_2-2=7}$$

Valor observado:

$$t_0 = \frac{2.8443 - 3.1063}{0.017527\sqrt{\frac{1}{4} + \frac{1}{5}}} = -22.28 \quad \text{valor} - \text{p} \approx 0$$

Pelo $I.C._{99\%}\left(\mu_1 - \mu_2\right) = [-0.3031; -0.2209]$

concluíamos que se rejeitava H_0 ao nível $\alpha = 1\%$.

Output do Excel para este teste:

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2.844275	3.1063
Variance	0.00041894	0.00022337
Observations	4	5
Pooled Variance	0.00030719	
Hyp. Mean Difference	0	
df	7	
t	-22.286055	
P(T<=t) one-tail	4.6302E-08	
t Critical one-tail	1.89457751	
P(T<=t) two-tail	9.2604E-08	
t Critical two-tail	2.36462256	

Teste relativo à igualdade das variâncias

(não faz parte do programa)

F-Test: Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	2.844275	3.1063
Variance	0.00041894	0.00022337
Observations	4	5
df	3	4
F	1.87551203	
P(F<=f) one-tail	0.27464723	
F Critical one-tail	6.59139232	

$H_0: \sigma_1^2 = \sigma_2^2$ contra $H_1: \sigma_1^2 \neq \sigma_2^2$

Estatística de teste:

$$F_0 = \frac{S_1^2}{S_2^2} \quad \text{sob } H_0 \quad F_0 \sim F_{n_1-1; n_2-1}$$

8.6 Testes de hipóteses para a variância de uma população normal

$X \sim N(\mu, \sigma^2)$ e $(X_1, \ldots, X_n)$ a.a.

Para testar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$

usa-se a estatística de teste

$$Q_0 = \frac{(n-1)S^2}{\sigma_0^2}$$

Quando H_0 é verdadeira $Q_0 \sim \chi^2_{n-1}$

Então, rejeita-se H_0 ao nível de significância α se

$$q_0 = \frac{(n-1)s^2}{\sigma_0^2} < a \quad \text{ou} \quad q_0 > b$$

com a: $P(Q_0 < a) = \dfrac{\alpha}{2}$ e b: $P(Q_0 > b) = \dfrac{\alpha}{2}$

8.7 Testes de hipóteses para uma proporção

$(X_1, \ldots, X_n)$ amostra aleatória de uma população muito grande ou infinita.

Seja $Y(\leq n)$ o número de observações desta amostra que pertencem a uma dada categoria de interesse.

Seja p a proporção de indivíduos na população que pertencem a essa categoria de interesse.

Exemplos:

População	Categoria
Peças	ser defeituosa
Eleitores	vota no partido X

O estimador pontual de p é $\hat{P} = \dfrac{Y}{n}$.

Já vimos que se n for grande

$$Z = \frac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \underset{a}{\sim} N(0,1)$$

Logo para testar $H_0 : p = p_0$ contra $H_1 : p \neq p_0$ (ou $H_1 : p < p_0$, ou $H_1 : p > p_0$) usa-se a estatística de teste

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}, \qquad \text{sob } H_0 \quad Z_0 \underset{a}{\sim} N(0,1)$$

Para $H_1 : p \neq p_0$, rejeita-se H_0 ao nível α se

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} < -a \text{ ou } z_0 > a$$

$$\left(a\text{: } P(Z > a) = \frac{\alpha}{2} \right)$$

Exemplo: População de eleitores portugueses. Sondagem (aleatória) a 1200 eleitores revelou que 683 tencionam votar no partido ABC. Entretanto o presidente do partido tinha afirmado "estou convencido que vamos obter mais de 50% dos votos". Concordamos com esta afirmação?

$\hat{p} = 683/1200 = 0.569$

Podemos testar $H_0 : p = 0.5$ contra $H_1 : p > 0.5$

Se rejeitarmos a hipótese nula (e isso é uma conclusão "forte") então a afirmação é corroborada pela sondagem.

$$z_0 = \frac{0.569 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{1200}}} = 4.79 \qquad \text{valor-p } = 0.000001$$

Como o valor-p é muito baixo rejeita-se H_0 para os níveis de significância usuais.

8.8 Teste do qui-quadrado de ajustamento

O objectivo é testar a hipótese de que as observações seguem uma determinada distribuição (discreta ou contínua, com ou sem parâmetros desconhecidos)

Exemplo: O lançamento de um dado 1000 vezes conduziu à seguinte tabela de frequências observadas (o_i)

x_i'	o_i
1	174
2	174
3	154
4	179
5	154
6	165
Total	1000

Será que os resultados obtidos sustentam a hipótese de que o "dado é perfeito"?

X - v.a. que representa o número de pontos obtido num lançamento

$$H_0: P(X=i)=\frac{1}{6},\ i=1,\ldots,6 \text{ ou } X\sim Unif.Disc.(1,\ldots,6)$$

H_1: negação de H_0

Quando H_0 é verdadeira sabemos calcular a probabilidade de cada valor (ou classe, em geral), que designamos por p_i, e o valor esperado para o número de observações em cada classe (abreviadamente, frequências esperadas),

$$E_i = np_i$$

onde n é a dimensão da amostra, neste caso $n=1000$

Vamos acrescentar essas duas colunas à tabela:

x_i'	o_i	p_i	$E_i = np_i$
1	174	1/6	166.67
2	174	1/6	166.67
3	154	1/6	166.67
4	179	1/6	166.67
5	154	1/6	166.67
6	165	1/6	166.67
Total	1000	1	1000.02

Mesmo quando H_0 é verdadeira não estamos à espera que as colunas o_i e E_i coincidam. É então necessário medir o afastamento entre o_i e E_i e saber até que ponto esse afastamento é razoável para H_0 verdadeira (se determinarmos que o afastamento é razoável não rejeitamos H_0, caso contrário rejeitamos H_0).

A variável que é usada para medir o afastamento é

$$X_0^2 = \sum_{i=1}^{k}\frac{\left(O_i - E_i\right)^2}{E_i} \qquad \text{(\underline{Estatística de teste})}$$

Pode mostrar-se que, quando H_0 é verdadeira,

$$X_0^2 \sim \chi^2_{k-\beta-1}$$

onde k é o nº. de classes (no exemplo, 6) e β é o nº. de parâmetros estimados (no exemplo, 0)

Deve rejeitar-se H_0 se o valor observado de X_0^2 for muito elevado, ou seja a região crítica do teste é da forma

$$R.C.: X_0^2 > a$$

onde a: $P\left(X_0^2 > a\right) = \alpha$

e α é o nível de significância do teste.

Tabela incluindo os cálculos para obter o valor observado de X_0^2:

x_i'	o_i	p_i	$E_i = np_i$	$\frac{(o_i - E_i)^2}{E_i}$
1	174	1/6	166.67	0.322
2	174	1/6	166.67	0.322
3	154	1/6	166.67	0.963
4	179	1/6	166.67	0.912
5	154	1/6	166.67	0.963
6	165	1/6	166.67	0.017
Total	1000	1	1000.02	3.499

O valor observado de X_0^2 é 3.499. Se fixarmos $\alpha = 0.05$, com $k - \beta - 1 = 5$, obtém-se $a = 11.07$.

Uma vez que 3.499 < 11.07, não se rejeita H_0 ao nível de significância de 5%.

Exemplo: Pensa-se que o número de defeitos por circuito, num certo tipo de circuitos, deve seguir uma distribuição de Poisson. De uma amostra (escolhida aleatoriamente) de 60 circuitos obtiveram-se os resultados seguintes:

Nº. de def.	o_i
0	32
1	15
2	9
3	4
Total	60

X - v.a. que representa o nº. de defeitos num circuito

H_0: $X \sim Poisson(\lambda)$ contra H_1: $X \sim outra\ dist.$

λ é desconhecido, então λ deve ser estimado (pelo método da máxima verosimilhança)

$$\hat{\lambda} = \bar{x} = \frac{32 \times 0 + 15 \times 1 + 9 \times 2 + 4 \times 3}{60} = 0.75$$

donde

$$\hat{p}_1 = \hat{P}(X = 0) = \frac{e^{-0.75} 0.75^0}{0!} = 0.472 \qquad e_1 = 28.32$$

$$\hat{p}_2 = \hat{P}(X = 1) = \frac{e^{-0.75} 0.75^1}{1!} = 0.354 \qquad e_2 = 21.24$$

$$\hat{p}_3 = \hat{P}(X = 2) = \frac{e^{-0.75} 0.75^2}{2!} = 0.133 \qquad e_3 = 7.98$$

$$\hat{p}_4 = \hat{P}(X \geq 3) = 1 - (\hat{p}_1 + \hat{p}_2 + \hat{p}_3) = 0.041$$

$$e_4 = 2.46$$

Deve ter-se $e_i \geq 5,\ \forall_i$, se para algum i $e_i < 5$, deve fazer-se um agrupamento de classes.

Obtém-se então a tabela final:

Nº. de def.	o_i	$\hat{p}_i$	$e_i = n\hat{p}_i$	$\frac{(o_i - e_i)^2}{e_i}$
0	32	0.472	28.32	0.478
1	15	0.354	21.24	1.833
≥ 2	13	0.174	10.44	0.628
Total	60	1.000	60.00	2.939

$$\begin{array}{l} k - \beta - 1 = 3 - 1 - 1 = 1 \\ \quad \alpha = 0.05 \end{array} \Rightarrow a = 3.841$$

Como 2.939 < 3.841, não se rejeita H_0 ao nível de significância de 5%.

Observações:

1) Para variáveis contínuas o procedimento é semelhante:

- As observações devem previamente ser agrupadas em classes (intervalos). Podem usar-se as regras para construção de histogramas e, à partida, classes de amplitude constante.
- p_i's são as probabilidades das classes.

2) É necessário n relativamente elevado para fazer este teste (pelo menos 5 observações por classe).

3) Existem outros testes que não requerem tantas observações (teste de Kolmogorov-Smirnov e papel de probabilidade) mas não fazem parte do programa.

8.9 Teste do qui-quadrado de independência em tabelas de contingência

O objectivo é testar a hipótese de que duas variáveis (discretas ou contínuas) são independentes. Para isso devemos ter observações relativas à ocorrência simultânea dos valores possíveis das duas variáveis. Essas observações organizam-se numa tabela de frequências a que se chama tabela de contingência.

Exemplo: Um estudo sobre a ocorrência de falhas numa certa componente electrónica revelou que podem ser considerados 4 tipos de falhas (A, B, C e D) e duas posições de montagem. Em 134 componentes seleccionadas aleatoriamente obtiveram-se as frequências absolutas registadas na tabela (de contingência) da página seguinte.

Será que o tipo de falha é independente da posição de montagem?

	Falha				
Montagem	A	B	C	D	Total
1	22	46	18	9	95
2	4	17	6	12	39
Total	26	63	24	21	134

Designamos por o_{ij}, (onde i se refere à linha e j à coluna) os valores do interior da tabela. Por $n_{i\bullet}$ os totais das colunas e por $n_{\bullet j}$ os totais das linhas.

Tabela genérica (com as mesmas dimensões):

	j				
i	1	2	3	4	$n_{i\bullet}$
1	o_{11}	o_{12}	o_{13}	o_{14}	$n_{1\bullet}$
2	o_{21}	o_{22}	o_{23}	o_{24}	$n_{2\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	$n_{\bullet 4}$	n

A hipótese nula (independência) pode ser escrita como:

$$H_0\colon P(X=i, Y=j) = P(X=i)P(Y=j) \quad \forall_{i,j}$$

ou $H_0\colon p_{ij} = p_{i\bullet} p_{\bullet j} \quad \forall_{i,j}$

Seguindo raciocínio semelhante ao usado no teste de ajustamento, precisamos de calcular a tabela de frequências esperadas sob a hipótese nula e compará-la com a de frequências observadas. Para isso é necessário primeiro estimar $p_{i\bullet}$ e $p_{\bullet j}$ $\forall_{i,j}$:

$$\hat{p}_{i\bullet} = \frac{n_{i\bullet}}{n} \qquad \hat{p}_{\bullet j} = \frac{n_{\bullet j}}{n},$$

donde se obtém

$$e_{ij} = n\hat{p}_{i\bullet}\hat{p}_{\bullet j} = n\frac{n_{i\bullet}}{n}\frac{n_{\bullet j}}{n} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

No exemplo em consideração obtém-se então a seguinte tabela de frequências esperadas:

	Falha			
Montagem	A	B	C	D
1	18.4	44.7	17.0	14.9
2	7.6	18.3	7.0	6.1

A variável que é usada para medir o afastamento (entre a tabela de frequências observadas e a tabela de frequências esperadas) é

$$X_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \qquad \text{(\underline{Estatística de teste})}$$

Pode mostrar-se que, quando H_0 é verdadeira,

$$X_0^2 \sim \chi^2_{(r-1)(s-1)}$$

com r=nº de linhas e s = nº de colunas da tabela.

Valor observado da estatística de teste no exemplo:

$$x_0^2 = \frac{(22-18.4)^2}{18.4} + \cdots + \frac{(12-6.1)^2}{6.1} = 10.78$$

Decisão: $\left((r-1)(s-1)=3\right)$

$$\alpha = 1\% \Rightarrow a\text{: } P\left(\chi_3^2 > a\right) = 0.99 \Leftrightarrow a = \chi^2_{3,0.99} = 11.34$$

$$\alpha = 2.5\% \Rightarrow a = \chi^2_{3,0.975} = 9.348$$

ou seja, $0.01 < \text{valor} - \text{p} < 0.025$

O resultado não é muito conclusivo, embora vá no sentido da não independência. Para ter um resultado mais convincente seria necessário repetir a experiência, eventualmente com mais observações.